

#2  
Docket No. 1359.1030/HJS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of

Chikako MATSUMOTO

Serial No.: Unassigned

Group Art Unit: Unassigned

Filed: November 28, 2000

Examiner: Unassigned

For: SPEECH DATA COMPRESSION/EXPANSION APPARATUS AND METHOD

**SUBMISSION OF CERTIFIED COPY OF PRIOR FOREIGN  
APPLICATION IN ACCORDANCE  
WITH THE REQUIREMENTS OF 37 C.F.R. §1.55**

*Honorable Commissioner of  
Patents and Trademarks  
Washington, D.C. 20231*

*Sir:*

In accordance with the provisions of 37 C.F.R. §1.55, the applicant(s) submit(s) herewith a certified copy of the following foreign application:

Japanese Patent Application No. 11-344615, filed December 3, 1999.

It is respectfully requested that the applicant(s) be given the benefit of the foreign filing date as evidenced by the certified papers attached hereto, in accordance with the requirements of 35 U.S.C. §119.

Respectfully submitted,

STAAS & HALSEY LLP

700 Eleventh Street, N.W.  
Washington, D.C. 20001  
(202) 434-1500

By: 

H. J. Staas

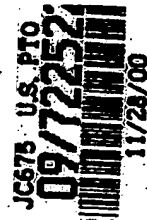
Registration No. 22,010

Date: November 28, 2000



日 本 国 特 許 庁

PATENT OFFICE  
JAPANESE GOVERNMENT



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日  
Date of Application:

1 9 9 9 年 1 2 月 3 日

出 願 番 号  
Application Number:

平成 1 1 年 特 許 願 第 3 4 4 6 1 5 号

出 願 人  
Applicant (s):

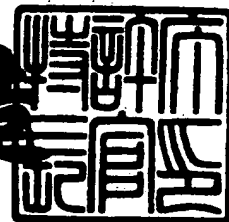
富士通株式会社

CERTIFIED COPY OF  
PRIORITY DOCUMENT

2 0 0 0 年 1 1 月 6 日

特 許 庁 長 官  
Commissioner,  
Patent Office

及 川 耕 造



出 証 番 号 出 証 特 2 0 0 0 - 3 0 9 1 9 4 9

【書類名】 特許願

【整理番号】 9995199

【提出日】 平成11年12月 3日

【あて先】 特許庁長官 殿

【国際特許分類】 G10L 9/00

【発明の名称】 音声データ圧縮・解凍装置及び方法

【請求項の数】 17

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

【氏名】 松本 智佳子

【特許出願人】

【識別番号】 000005223

【氏名又は名称】 富士通株式会社

【代理人】

【識別番号】 100095555

【弁理士】

【氏名又は名称】 池内 寛幸

【電話番号】 06-6361-9334

【手数料の表示】

【予納台帳番号】 012162

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9803089

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声データ圧縮・解凍装置及び方法

【特許請求の範囲】

【請求項 1】 波形データを含む音声データを既存の音声波形辞書から抽出して入力する辞書データ入力部と、

前記波形データの中で音声合成に用いる部分を特定し、その前後で圧縮する開始点及び終了点を設定する圧縮位置決定部と、

設定した圧縮区間について前記波形データを圧縮する辞書データ圧縮部と、

圧縮された前記波形データを解凍する辞書データ解凍部とを含み、

圧縮された前記波形データの解凍結果の品質が最も高かった前記圧縮する開始点及び終了点で指定される前記圧縮区間を圧縮・解凍位置として確定し、音声合成に用いる前記波形データとして前記圧縮された波形データと前記圧縮する開始点及び終了点をデータベースに登録することを特徴とした音声データ圧縮・解凍装置。

【請求項 2】 前記圧縮位置決定部において、前記波形データの中で音声合成に用いる部分を特定し、その前後で圧縮する開始点及び終了点を仮設定し、

仮設定した圧縮区間について前記波形データを圧縮する辞書データ圧縮部と、

圧縮された前記波形データを解凍する辞書データ解凍部と、

解凍された前記波形データについて S/N 比を計算する S/N 比計算部とを含み、

前記 S/N 比が最も高かった前記圧縮する開始点及び終了点で指定される前記圧縮区間を圧縮・解凍位置として確定し、音声合成に用いる前記波形データとして前記圧縮された波形データをデータベースに登録する請求項 1 記載の音声データ圧縮・解凍装置。

【請求項 3】 音声合成に用いる前記波形データとしてデータベースに登録された圧縮された前記波形データの前後で解凍する開始点及び終了点を設定する解凍位置決定部をさらに含み、

前記辞書データ解凍部において、設定した解凍区間について前記波形データを解凍する請求項 1 記載の音声データ圧縮・解凍装置。

【請求項 4】 前記圧縮位置決定部において、ピッチ単位で圧縮する開始点及び終了点を定める請求項 1 記載の音声データ圧縮・解凍装置。

【請求項 5】 前記圧縮位置決定部において、フレーム単位で圧縮する開始点及び終了点を定める請求項 1 記載の音声データ圧縮・解凍装置。

【請求項 6】 データベースに保存されている請求項 1 又は 5 記載の装置で圧縮された前記波形データを解凍することを特徴とする音声データ解凍装置。

【請求項 7】 波形データを含む音声データを既存の音声波形辞書から抽出して入力する辞書データ入力部と、

前記波形データの中で音声合成に用いる部分を特定し、その部分を含む圧縮位置を決定する圧縮位置決定部と、

設定した圧縮区間について前記波形データを圧縮する辞書データ圧縮部と、

圧縮された前記波形データの前後で解凍する開始点と終了点を設定する解凍位置決定部と、

設定した解凍区間について圧縮された前記波形データを解凍する辞書データ解凍部とを含み、

圧縮された前記波形データの解凍結果の品質が最も高かった前記解凍する開始点及び終了点で指定される前記解凍区間を解凍位置として確定し、音声合成に用いる前記波形データとして前記圧縮された波形データと前記解凍する開始点及び終了点をデータベースに登録することを特徴とした音声データ圧縮・解凍装置。

【請求項 8】 データベースに保存されている請求項 7 記載の装置で前記解凍区間の確定した前記波形データを解凍することを特徴とした音声データ解凍装置

【請求項 9】 前記解凍位置決定部において、圧縮された前記波形データの前後で解凍する開始点と終了点を仮設定し、

仮設定した解凍区間について圧縮された前記波形データを解凍する辞書データ解凍部と、

解凍された前記波形データについて S/N 比を計算する S/N 比計算部とを含み、

前記 S/N 比が最も高かった前記解凍する開始点及び終了点で指定される前記

解凍区間を解凍位置として確定する請求項 7 記載の音声データ圧縮・解凍装置。

【請求項 1 0】 前記解凍位置決定部において、ピッチ単位で解凍する開始点及び終了点を定める請求項 7 記載の音声データ圧縮・解凍装置。

【請求項 1 1】 前記解凍位置決定部において、ビット詰めするバイト数と開始点から解凍する終了点を定める請求項 7 記載の音声データ圧縮・解凍装置。

【請求項 1 2】 波形データを含む音声データを既存の音声波形辞書から抽出して入力する工程と、

前記波形データの中で音声合成に用いる部分を特定し、その前後で圧縮する開始点及び終了点を設定する工程と、

設定した圧縮区間について前記波形データを圧縮する工程と、

圧縮された前記波形データを解凍する工程とを含み、

圧縮された前記波形データの解凍結果の品質が最も高かった前記圧縮する開始点及び終了点で指定される前記圧縮区間を圧縮・解凍位置として確定し、音声合成に用いる前記波形データとして前記圧縮された波形データと前記圧縮する開始点及び終了点をデータベースに登録することを特徴とする音声データ圧縮・解凍方法。

【請求項 1 3】 波形データを含む音声データを既存の音声波形辞書から抽出して入力する工程と、

前記波形データの中で音声合成に用いる部分を特定し、その部分を含む圧縮位置を決定する工程と、

設定した圧縮区間について前記波形データを圧縮する工程と、

圧縮された前記波形データの前後で解凍する開始点と終了点を設定する工程と

、  
設定した解凍区間について圧縮された前記波形データを解凍する工程とを含み

、  
圧縮された前記波形データの解凍結果の品質が最も高かった前記解凍する開始点及び終了点で指定される前記解凍区間を解凍位置として確定し、音声合成に用いる前記波形データとして前記圧縮された波形データと前記解凍する開始点及び終了点をデータベースに登録することを特徴とする音声データ圧縮・解凍方法。

【請求項 1 4】 データベースに保存されている請求項 1 又は 5 記載の装置で圧縮された前記波形データを解凍することを特徴とする音声データ解凍方式。

【請求項 1 5】 データベースに保存されている請求項 7 記載の装置で前記解凍区間が確定された前記波形データを解凍することを特徴とする音声データ解凍方式。

【請求項 1 6】 波形データを含む音声データを既存の音声波形辞書から抽出して入力するステップと、

前記波形データの中で音声合成に用いる部分を特定し、その前後で圧縮する開始点及び終了点を設定するステップと、

設定した圧縮区間について前記波形データを圧縮するステップと、

圧縮された前記波形データを解凍するステップとを含み、

圧縮された前記波形データの解凍結果の品質が最も高かった前記圧縮する開始点及び終了点で指定される前記圧縮区間を圧縮・解凍位置として確定し、音声合成に用いる前記波形データとして前記圧縮された波形データと前記圧縮する開始点及び終了点をデータベースに登録することを特徴とするコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項 1 7】 波形データを含む音声データを既存の音声波形辞書から抽出して入力するステップと、

前記波形データの中で音声合成に用いる部分を特定し、その部分を含む圧縮位置を決定するステップと、

設定した圧縮区間について前記波形データを圧縮するステップと、

圧縮された前記波形データの前後で解凍する開始点と終了点を設定するステップと、

設定した解凍区間について圧縮された前記波形データを解凍するステップとを含み、

圧縮された前記波形データの解凍結果の品質が最も高かった前記解凍する開始点及び終了点で指定される前記解凍区間を解凍位置として確定し、音声合成に用いる前記波形データとして前記圧縮された波形データと前記解凍する開始点及び終了点をデータベースに登録することを特徴とするコンピュータに実行させるプ

ログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声合成に使用される音声波形データからなる波形辞書データを圧縮して圧縮辞書を作成する圧縮装置及び圧縮されたデータを解凍する解凍装置に関する。

【0002】

【従来の技術】

昨今の急速なコンピュータ技術の発達によって、従来特定分野に利用が限定されてきた音声合成技術が、様々な分野で適用できるようになってきた。それに伴い、音声合成において、さらなる高品質な音声による再生が求められるようになってきている。

【0003】

高品質な音声合成を実現するためには、比較的大容量のデータである音の波形データを大量に準備しておく必要があり、ディスク等の記憶装置といった計算機資源を大量に消費することになる。したがって、かかる音の波形データを圧縮する方法については、様々な手法が考えられている。

【0004】

例えば図1には、従来から良く用いられている圧縮・解凍装置の原理図を示している。図1において、11は辞書データ入力部を、12は辞書データ圧縮部を、13は圧縮辞書データ保存部を、14は音声辞書データベースを、15は辞書データ解凍部を、16は解凍波形データ出力部を、それぞれ示す。

【0005】

図1では、辞書データは、波形データ111、音素ラベル112、ピッチ情報113とで構成されているが、かかる従来の圧縮・解凍装置においては波形データ111のみが圧縮・解凍の対象となっている。したがって、辞書データ圧縮部12では入力された波形データ111について圧縮処理を行い、圧縮辞書データ保存部13によって音声辞書データベース14に保存される。



## 【0 0 0 6】

また、音声辞書データベース 1 4 に保存されている圧縮された波形データについては、音声合成再生時に辞書データ解凍部 1 5 において解凍され、解凍波形データ出力部 1 6 において再生される。

## 【0 0 0 7】

## 【発明が解決しようとする課題】

しかし、かかる圧縮・解凍方法では、従来の波形データをそのまま圧縮対象としていることから、元になる辞書における波形データが音素単位ではなくコーパス単位であった場合等には、音声合成で用いる対象となる音素や音節がコーパス内のどの部分に相当するのか等の判断が困難であるとともに、コーパス単位で圧縮されたデータのすべてを解凍する必要があることから、解凍自体にも相当の時間を要し、リアルタイムで音声合成することが困難であるという問題点があった。

## 【0 0 0 8】

さらに、圧縮された音声波形データを解凍して音声合成する場合において、音声合成の立ち上がり部分で S/N 比が悪くなるが多いために高品質で再生することが困難であるという問題点もあった。

## 【0 0 0 9】

本発明では、上記問題点を解消すべく、波形データにおける圧縮位置や解凍位置を是正して、音声合成のリアルタイム性を確保するとともに、高品質な音声合成を実現する音声データ圧縮・解凍装置及び方法を提供することを目的とする。

## 【0 0 1 0】

## 【課題を解決するための手段】

上記目的を達成するために本発明にかかる音声データ圧縮・解凍装置は、波形データを含む音声データを既存の音声波形辞書から抽出して入力する辞書データ入力部と、波形データの中で音声合成に用いる部分を特定し、その前後で圧縮する開始点及び終了点を設定する圧縮位置決定部と、設定した圧縮区間について波形データを圧縮する辞書データ圧縮部と、圧縮された波形データを解凍する辞書データ解凍部とを含み、圧縮された波形データの解凍結果の品質が最も高かった

圧縮する開始点及び終了点で指定される圧縮区間を圧縮・解凍位置として確定し、音声合成に用いる波形データとして圧縮された波形データと圧縮する開始点及び終了点をデータベースに登録することを特徴とする。

## 【 0 0 1 1 】

かかる構成により、波形データにおける圧縮位置を、任意に決めることができ、圧縮すべき波形データの容量を必要最小限に抑制することができることから、解凍時間を短縮することができ、音声合成時のリアルタイム性を確保することが可能となる。

## 【 0 0 1 2 】

また、本発明にかかる音声データ圧縮・解凍装置は、圧縮位置決定部において、波形データの中で音声合成に用いる部分を特定し、その前後で圧縮する開始点及び終了点を仮設定し、仮設定した圧縮区間について波形データを圧縮する辞書データ圧縮部と、圧縮された波形データを解凍する辞書データ解凍部と、解凍された波形データについて S/N 比を計算する S/N 比計算部とを含み、S/N 比が最も高かった圧縮する開始点及び終了点で指定される圧縮区間を圧縮・解凍位置として確定し、音声合成に用いる波形データとして圧縮された波形データをデータベースに登録することが好ましい。

## 【 0 0 1 3 】

かかる構成により、波形データにおける圧縮位置を、音声合成時の S/N 比が高い位置で決定することができ、品質の高い音声合成を行うことが可能となるとともに、圧縮すべき波形データを必要最小限に抑制することができることから、解凍時間を短縮することができ、音声合成のリアルタイム性を確保することが可能となる。

## 【 0 0 1 4 】

また、本発明にかかる音声データ圧縮・解凍装置は、音声合成に用いる波形データとしてデータベースに登録された圧縮された波形データの前後で解凍する開始点及び終了点を設定する解凍位置決定部をさらに含み、辞書データ解凍部において、設定した解凍区間について波形データを解凍することが好ましい。波形データにおける解凍位置を任意に決めることができ、品質の高い音声合成を行うこ

とが可能となるからである。

【0015】

また、本発明にかかる音声データ圧縮・解凍装置は、圧縮位置決定部において、ピッチ単位で圧縮する開始点及び終了点を定めることが好ましい。さらに、本発明にかかる音声データ圧縮・解凍装置は、圧縮位置決定部において、フレーム単位で圧縮する開始点及び終了点を定めることが好ましい。圧縮する開始点及び終了点を容易に特定することができるからである。

【0016】

次に、上記目的を達成するために本発明にかかる音声データ解凍装置は、データベースに保存されている上述した本発明にかかる音声データ圧縮・解凍装置によって圧縮された波形データを解凍することを特徴とする。

【0017】

かかる構成により、圧縮された波形データを保存したデータベースを用いることで、母集団の大きな波形データを保持することが可能となり、その中からより適切な波形データを選択して解凍することができる。したがって、本発明にかかる音声データ解凍装置を用いることによって、より質の高い音声合成装置を構成することが可能となる。

【0018】

次に、上記目的を達成するために本発明にかかる音声データ圧縮・解凍装置は、波形データを含む音声データを既存の音声波形辞書から抽出して入力する辞書データ入力部と、波形データの中で音声合成に用いる部分を特定し、その部分を含む圧縮位置を決定する圧縮位置決定部と、設定した圧縮区間について波形データを圧縮する辞書データ圧縮部と、圧縮された波形データの前後で解凍する開始点と終了点を設定する解凍位置決定部と、設定した解凍区間について圧縮された波形データを解凍する辞書データ解凍部とを含み、圧縮された波形データの解凍結果の品質が最も高かった解凍する開始点及び終了点で指定される解凍区間を解凍位置として確定し、音声合成に用いる波形データとして圧縮された波形データと解凍する開始点及び終了点をデータベースに登録することを特徴とする。

【0019】

かかる構成により、波形データにおける解凍位置を任意に決めることができ、解凍すべき波形データの容量を必要最小限に抑制することができることから、解凍時間を短縮することができ、音声合成のリアルタイム性を確保することが可能となる。

#### 【 0 0 2 0 】

次に、上記目的を達成するために本発明にかかる音声データ解凍装置は、データベースに保存されている上述した本発明にかかる音声データ圧縮・解凍装置によって解凍区間が確定した波形データを解凍することを特徴とする。

#### 【 0 0 2 1 】

かかる構成により、圧縮された波形データを保存したデータベースを用いることで、母集団の大きな波形データを保持することが可能となり、その中からより適切な波形データを選択して解凍することができるとともに、より解凍品質の高い波形データを用いることができることから、本発明にかかる音声データ解凍装置を用いることによって、より質の高い音声合成装置を構成することが可能となる。

#### 【 0 0 2 2 】

また、本発明にかかる音声データ圧縮・解凍装置は、解凍位置決定部において、圧縮された波形データの前後で解凍する開始点と終了点を仮設定し、仮設定した解凍区間について圧縮された波形データを解凍する辞書データ解凍部と、解凍された波形データについて  $S/N$  比を計算する  $S/N$  比計算部とを含み、 $S/N$  比が最も高かった解凍する開始点及び終了点で指定される解凍区間を解凍位置として確定することが好ましい。圧縮された波形データにおける解凍位置を、音声合成時の  $S/N$  比が高い位置で決定することができ、品質の高い音声合成を行うことが可能となるからである。

#### 【 0 0 2 3 】

また、本発明にかかる音声データ圧縮・解凍装置は、解凍位置決定部において、ピッチ単位で解凍する開始点及び終了点を定めることが好ましい。さらに、本発明にかかる音声データ圧縮・解凍装置は、解凍位置決定部において、ビット詰めするバイト数と開始点から解凍する終了点を定めることが好ましい。圧縮され

た波形データを解凍する開始点及び終了点を容易に特定することができるからである。

## 【 0 0 2 4 】

次に、上記目的を達成するために本発明にかかる音声データ解凍方式は、データベースに保存されている上述した本発明にかかる音声データ圧縮・解凍装置によって圧縮された波形データを解凍することを特徴とする。

## 【 0 0 2 5 】

かかる構成により、圧縮された波形データを保存したデータベースを用いることで、母集団の大きな波形データを保持することが可能となり、その中からより適切な波形データを選択して解凍することができる。したがって、本発明にかかる音声データ解凍装置を用いることによって、より質の高い音声合成装置を構成することが可能となる。

## 【 0 0 2 6 】

次に、上記目的を達成するために本発明にかかる音声データ解凍方式は、データベースに保存されている上述した本発明にかかる音声データ圧縮・解凍装置によって解凍区間が確定された波形データを解凍することを特徴とする。

## 【 0 0 2 7 】

かかる構成により、圧縮された波形データを保存したデータベースを用いることで、母集団の大きな波形データを保持することが可能となり、その中からより適切な波形データを選択して解凍することができるとともに、より解凍品質の高い波形データを用いることができることから、本発明にかかる音声データ解凍装置を用いることによって、より質の高い音声合成装置を構成することが可能となる。

## 【 0 0 2 8 】

また、本発明は、上記のような音声データ圧縮・解凍装置の機能をコンピュータの処理ステップとして実行するソフトウェアを特徴とするものであり、具体的には、波形データを含む音声データを既存の音声波形辞書から抽出して入力する工程と、波形データの中で音声合成に用いる部分を特定し、その前後で圧縮する開始点及び終了点を設定する工程と、設定した圧縮区間について波形データを圧

縮する工程と、圧縮された波形データを解凍する工程とを含み、圧縮された波形データの解凍結果の品質が最も高かった圧縮する開始点及び終了点で指定される圧縮区間を圧縮・解凍位置として確定し、音声合成に用いる波形データとして圧縮された波形データと圧縮する開始点及び終了点をデータベースに登録する方法並びにそのような工程をプログラムとして記録したコンピュータ読み取り可能な記録媒体であることを特徴とする。

## 【 0 0 2 9 】

かかる構成により、コンピュータ上へ当該プログラムをロードさせ実行することで、波形データにおける圧縮位置を、任意に決めることができ、圧縮すべき波形データの容量を必要最小限に抑制することができることから、解凍時間を短縮することができ、音声合成のリアルタイム性を確保することができる音声データ圧縮・解凍装置を実現することが可能となる。

## 【 0 0 3 0 】

また、本発明は、上記のような音声データ圧縮・解凍装置の機能をコンピュータの処理ステップとして実行するソフトウェアを特徴とするものであり、具体的には、波形データを含む音声データを既存の音声波形辞書から抽出して入力する工程と、波形データの中で音声合成に用いる部分を特定し、その部分を含む圧縮位置を決定する工程と、設定した圧縮区間について波形データを圧縮する工程と、圧縮された波形データの前後で解凍する開始点と終了点を設定する工程と、設定した解凍区間について圧縮された波形データを解凍する工程とを含み、圧縮された波形データの解凍結果の品質が最も高かった解凍する開始点及び終了点で指定される解凍区間を解凍位置として確定し、音声合成に用いる波形データとして圧縮された波形データと解凍する開始点及び終了点をデータベースに登録する方法並びにそのような工程をプログラムとして記録したコンピュータ読み取り可能な記録媒体であることを特徴とする。

## 【 0 0 3 1 】

かかる構成により、コンピュータ上へ当該プログラムをロードさせ実行することで、母集団の大きな波形データの中からより適切な波形データを選択することができるので、より質の高い音声合成装置を実現することが可能となる。

## 【 0 0 3 2 】

## 【発明の実施の形態】

以下、本発明の実施の形態にかかる音声データ圧縮・解凍装置について、図面を参照しながら説明する。図 2 は本発明の実施の形態にかかる音声データ圧縮・解凍装置の原理構成図を示す。図 2 において、2 1 は圧縮辞書データ保存部を、2 2 は圧縮位置決定部を、2 3 は解凍位置決定部を、2 4 は S / N 比計算部を、それぞれ示す。

## 【 0 0 3 3 】

まず、図 2 において、辞書データは図 1 の従来の場合と同様に波形データ 1 1 1、音素ラベル 1 1 2、ピッチ情報 1 1 3 とで構成されている。本実施の形態においては従来の圧縮・解凍装置と同様に、波形データ 1 1 1 のみが圧縮・解凍の対象となっている。しかし、波形データ 1 1 1 のすべてについて圧縮の対象とするのではなく、圧縮の対象となる区間、すなわち圧縮処理の開始点と終了点を設定し、当該区間のみを圧縮の対象とするものである。たがって、辞書データ圧縮部 1 2 では入力された波形データ 1 1 1 のみならず、圧縮位置の決定に必要な情報として音素ラベル 1 1 2、ピッチ情報 1 1 3 についても、圧縮辞書データ保存部 2 1 によって音声辞書データベース 1 4 に保存される。

## 【 0 0 3 4 】

上述した圧縮位置の決定方法としては、様々な方法が考えられる。まず、圧縮処理の開始点と終了点を変更しつつ解凍を行い、各々の場合に測定される S / N 比に基づいて音素や音節単位で最も S / N 比の高かった区間を圧縮区間とする方法が考えられる。この場合、圧縮位置は一度では確定できず、圧縮位置決定部 2 2 において、図 3 に示すような処理を行うことによって決定する。図 3 は本発明の実施の形態にかかる音声データ圧縮・解凍装置における波形データ圧縮時の概念説明図である。図 3 において、3 1 が圧縮の対象となる波形データを、3 2 が圧縮の対象となる波形データの前後に有する付加データを、それぞれ示す。

## 【 0 0 3 5 】

まず図 3 では、元の波形データ全体を示す ( a ) において、音声合成に用いる波形データ 3 1 の開始点 3 3 及び終了点 3 4 を定める。そして、圧縮の対象とな

る波形データ 31 をそのまま圧縮したのでは、解凍時に音声の立ち上がり部分において  $S/N$  比を高く維持することが困難であることから、圧縮の対象となる波形データ 31 の前後において圧縮時の開始点及び終了点を仮に設定する。すなわち、適当なデータ長を有する付随データ 32 を音声合成に用いる波形データ 31 の前後に含めて、圧縮の開始点 35 及び圧縮の終了点 36 を仮設定することになる。なお、かかる付随データ 32 のデータ長を定める単位としては、フレーム単位であっても良いし、コーパス等のサンプル単位やピッチ単位であっても良い。

## 【0036】

そして、(b) に示すように付随データ 32 も含めて波形データ 31 を圧縮してから、辞書データ解凍部 15 で (c) に示すように波形データ 31 を解凍する。解凍された波形データ 31 は、付随データ 32 の先端部においてはノイズの影響で  $S/N$  比が低いが、音声合成に用いる部分の波形データ 31 については  $S/N$  比を高く維持した状態で取得することができる。したがって、音声合成に用いる部分の波形データ区間 37 を残して付随データ 32 を削除することで、 $S/N$  比の高い解凍波形データを得ることが可能となる。

## 【0037】

そして、解凍位置決定部 23 では、得られた解凍波形データにおける音声合成に用いる部分と解凍する区間の開始点及び終了点を一致させるとともに、解凍波形データと元の波形データとの  $S/N$  比を  $S/N$  比計算部 24 で算出して、圧縮位置決定部 22 へその計算結果を送る。

## 【0038】

圧縮位置決定部 22 では、上述したような処理を圧縮時の開始点と終了点を変更しながら繰り返し  $S/N$  比の計算結果を取得し、 $S/N$  比の計算結果のうちで最も  $S/N$  比の高い圧縮位置を求め、圧縮位置情報 144 として保存する。

## 【0039】

また、フレーム単位で圧縮区間の終了点を決定する方法も考えられる。この場合、圧縮位置決定部 22 では、辞書データ圧縮部 12 におけるフレームの単位に基づいて圧縮区間の終了点を決定する。

## 【0040】



さらに、元データから無音区間を取り除いて有音区間だけとし、有音区間を圧縮区間とする方法も考えられる。この場合、圧縮位置決定部 22 では、音素ラベル 112 とピッチ情報 113 から無音区間を抽出・削除し、有音区間を圧縮区間として決定する。

#### 【0041】

また、圧縮位置の仮設定処理を排するため、元の波形データの単位そのもの、例えばコーパス単位で波形データが取得されている場合にはコーパス単位で圧縮する方法や、等間隔で波形データを区切る方法、あるいは辞書データの音素ラベル 112 とピッチ情報 113 に基づいて、数ピッチ前を圧縮区間の開始点とする方法等も考えられる。

#### 【0042】

これらの方法においては、圧縮位置決定部 22 において一度で圧縮位置を決定することができることから、圧縮位置決定部 22 で決定された圧縮位置の開始点及び終了点を圧縮波形データ 141 として音声辞書データベース 14 に保存する。

#### 【0043】

そして、音声合成に用いる波形データが圧縮された波形データの一部である場合には、解凍位置決定部 23 において解凍時の区間を決定して、解凍位置情報 145 に保存しておく。

#### 【0044】

ここで、解凍位置の決定方法には大きく 3 つの方法が考えられる。一つには、解凍区間の開始点と終了点を変更しつつ解凍を行い、各々の場合に測定される S/N 比に基づいて音素や音節単位で最も S/N 比の高かった区間を解凍区間とする方法、あるいは音素ラベルとピッチ情報から、自動的に解凍時の開始点を数ピッチ前に設定する方法、さらに解凍結果より判明するビット詰めするバイト数と開始点に基づいて、解凍区間の終了点を自動的に計算して解凍区間とする方法である。

#### 【0045】

まず、解凍区間の開始点と終了点を変更しつつ解凍を行い、各々の場合に測定

される  $S/N$  比に基づいて音素や音節単位で最も  $S/N$  比の高かった区間を解凍区間とする方法の場合には、解凍位置は一度では確定できず、解凍位置決定部 23 において、図 4 に示すような処理を行うことによって決定する。図 4 は本発明の実施の形態にかかる音声データ圧縮・解凍装置における波形データ解凍時の概念説明図である。図 4 において、41 が圧縮の対象となる波形データを、42 が圧縮された波形データの前後に有する付加データを、それぞれ示す。

## 【0046】

図 4 において、音声合成に用いるべき波形データは、図 4 (b) に示すような圧縮された状態で音声辞書データベース 14 に登録されている。かかる圧縮された波形データをそのまま解凍すれば、元の波形データ全体を示す図 4 (a) のようになる。したがって、音声合成に用いる波形データ 41 の開始点 43 及び終了点 44 では、解凍時に  $S/N$  比が低くなる蓋然性が高い。

## 【0047】

そこで、解凍時に音声合成に用いる波形データにノイズを拾わせないために、圧縮されている波形データ 48 の前後に適当なデータ長を有する付加データ 42 を付加し、解凍の開始点 45 及び解凍の終了点 46 を仮に設定する。かかる付加データのデータ長を定める単位としては、フレーム単位であっても良いし、コーパス等のサンプル単位やピッチ単位であっても良い。

## 【0048】

そして、辞書データ解凍部 15 で、図 4 (c) に示すように圧縮データ 49 を解凍する。解凍された波形データ 47 については、付加データ 42 の先端部においてはノイズの影響で  $S/N$  比が低いが、音声合成に用いる波形データ 47 については  $S/N$  比を高く維持した状態で取得することができる。したがって、音声合成に用いる波形データ区間 47 を残して付加データを削除することで、解凍波形データを得ることが可能となる。

## 【0049】

そして、解凍位置決定部 23 では、得られた解凍波形データにおける音声合成に用いる部分と解凍する区間の開始点及び終了点を一致させるとともに、解凍波形データと元の波形データとの  $S/N$  比を  $S/N$  比計算部 24 で算出して、解凍

位置決定部 23 へその計算結果を送る。

【0050】

解凍位置決定部 23 では、解凍時の開始点と終了点を変更しながら  $S/N$  比の計算結果を取得し、最も  $S/N$  比の高い解凍位置を求めることで、解凍位置情報として保存する。

【0051】

また、音素ラベルとピッチ情報から、自動的に解凍時の開始点を数ピッチ前に設定する方法の場合には、解凍位置決定部 23 において解凍位置を一度で確定することができる。

【0052】

さらには、圧縮結果をビット詰めするバイト数と開始点から終了点を自動的に計算して解凍区間とする方法の場合には、解凍位置決定部 23 において、解凍時のビット詰めするバイト数と開始点から、終了時点を自動的に算出するものであり、求められた区間を解凍区間として解凍位置情報として保存する。

【0053】

また、音声辞書データベース 14 に保存されている圧縮された波形データについては、音声合成再生時に辞書データ解凍部 15 において解凍され、解凍波形データ出力部 16 において再生される。すなわち、図 5 に示すように、音声合成部 51 を配することで、音節ごとに合成音声による再生が可能となる。

【0054】

より具体的には、以下のようなになる。図 6 は本発明の一実施例にかかる音声データ圧縮・解凍装置の構成図である。まず、圧縮位置決定部 22 及び解凍位置決定部 23 は図 6 に示すように構成されている。すなわち、圧縮位置決定部 22 において、221 は無音区間削除部を、222 は有音区間波形生成部を、223 は圧縮区間設定部を、それぞれ示す。また、解凍位置決定部 23 において、231 は音節抽出部を、232 は音節波形区間抽出部を、233 は解凍区間設定部を、234 は解凍区間・ $S/N$  比保存部を、それぞれ示す。

【0055】

まず、「私は犬を飼っています。」というコーパスの波形データが音声辞書デ

ータベース 1 4 に保存されているものとする、無音区間削除部 2 2 1 において音素ラベル 1 1 2 及びピッチ情報 1 1 3 に基づいて波形データ 1 1 1 の無音区間を抽出して、削除する。そして、有音区間波形生成部 2 2 2 において、有音部のみからなる波形を生成し、波形データ 1 1 1 として保存する。

【0 0 5 6】

圧縮区間設定部 2 2 3 では、コーパスの最初から最後まで有音区間全部を指定し、その開始点と終了点を圧縮位置情報 1 4 4 に保存する。そして、「私は犬を飼っています。」というコーパス内の有音部分の波形データを圧縮し、結果を圧縮波形データ 1 4 1 に保存する。

【0 0 5 7】

辞書データ圧縮部 1 2 では、「私は犬を飼っています。」というコーパス内の有音部分の波形データを圧縮して、結果を圧縮波形データ 1 4 1 に保存する。保存されている圧縮された波形データに関する新たな音素ラベル及びピッチ情報についても、音素ラベル 1 4 2、ピッチ情報 1 4 3 として音声辞書データベース 1 4 に保存される。

【0 0 5 8】

また、解凍区間を設定する際には、音素抽出部 2 3 1 で「私は犬を飼っています。」というコーパス内の音節部を抽出する。すなわち、「私は」、「犬を」、「飼って」、「います」という 4 つの音節部が抽出される。

【0 0 5 9】

そして、抽出された各音節について、音節波形区間抽出部 2 3 2 において、各音節ごとの圧縮前の波形データ 1 1 1 における開始位置と終了位置を検出する。解凍区間設定部 2 3 3 では、圧縮波形データ 1 4 1 における開始位置と終了位置を、各音節ごとの圧縮前の波形データ 1 1 1 における開始位置と終了位置に基づいて仮設定する。

【0 0 6 0】

設定方法としては、種々の方法が考えられるが、例えば解凍時の開始位置又は終了位置を、必要とされる圧縮前の波形データ 1 1 1 における開始位置又は終了位置の 1 から数フレーム前後にする方法や、解凍時の開始位置又は終了位置を、

必要とされる圧縮前の波形データ 1 1 1 における開始位置又は終了位置の 1 から数サンプル前後にする方法、あるいは解凍時の開始位置又は終了位置を、必要とされる圧縮前の波形データ 1 1 1 における開始位置又は終了位置の 1 から数ピッチ前後にする方法等が考えられる。

## 【 0 0 6 1 】

辞書データ解凍部 1 5 では、解凍区間設定部 2 3 3 で仮設定された解凍区間について実際に解凍し、S/N 比計算部 2 4 で S/N 比を計算して、解凍区間・S/N 比保存部 2 3 4 に保存される。そして、解凍区間・S/N 比保存部 2 3 4 に保存されているデータの中で最も S/N 比の高い区間データを解凍区間として、当該区間データの開始位置及び終了位置を解凍位置保存部 1 4 5 に保存する。

## 【 0 0 6 2 】

実際の解凍時には、解凍したい対象となる音節が入力されると、辞書データ解凍部 1 5 においては、解凍位置保存部 1 4 5 に保存されている区間データに基づいて解凍を行う。そして、解凍された波形データについては、必要な部分だけカットして使用される。

## 【 0 0 6 3 】

また、図 7 は本発明の他の実施例にかかる音声データ圧縮・解凍装置の構成図である。図 6 とは圧縮位置決定部 2 2 の構成を除いて同一の構成である。したがって、解凍位置決定部 2 3 についての詳細な記述は省略している。圧縮位置決定部 2 2 において、2 2 4 は音節抽出部を、2 2 5 は圧縮区間・S/N 比保存部を、それぞれ示す。

## 【 0 0 6 4 】

図 6 の場合と同様に、「私は犬を飼っています。」というコーパスの波形データが音声辞書データベース 1 4 に保存されているものとする。無音区間削除部 2 2 1 において音素ラベル 1 1 2 及びピッチ情報 1 1 3 に基づいて波形データ 1 1 1 の無音区間を抽出して、削除する。そして、有音区間波形生成部 2 2 2 において、有音部のみからなる波形を生成し、波形データ 1 1 1 として保存する。

## 【 0 0 6 5 】

音節抽出部 2 2 4 では、「私は犬を飼っています。」というコーパス内の音節

部を抽出する。すなわち、「私は」、「犬を」、「飼って」、「います」という 4 つの音節部が抽出される。

【0066】

圧縮区間設定部 2 2 3 では、抽出された各音節、例えば「犬を」における圧縮前の波形データの開始位置と終了位置の前後に図 3 に示すように付加データを付加して、圧縮区間を仮に設定し、当該圧縮区間内のデータについて辞書データ圧縮部 1 2 によって圧縮する。圧縮する方法については、前述の通りである。

【0067】

圧縮されたデータは、一旦辞書データ解凍部 1 5 によって解凍され、解凍波形データ出力部 1 6 によって出力される解凍された波形データと、圧縮前の波形データ 1 1 1 との  $S/N$  比を  $S/N$  比計算部 2 4 で計算し、圧縮区間・ $S/N$  比保存部 2 2 5 に圧縮区間の開始点及び終了点とともに保存される。

【0068】

そして、圧縮区間・ $S/N$  比保存部 2 2 5 に保存されているデータのうち、最も  $S/N$  比の高い区間データを解凍区間として、当該区間データの開始位置及び終了位置を解凍位置保存部 1 4 5 に保存する。

【0069】

実際の解凍時には、解凍したい対象となる音節が入力されると、辞書データ解凍部 1 5 においては、解凍位置保存部 1 4 5 に保存されている区間データに基づいて解凍を行う。そして、解凍された波形データについては、必要な部分だけカットして使用される。

【0070】

以上のように本実施の形態によれば、波形データにおける圧縮位置や解凍位置を、音声合成時の  $S/N$  比が高くなる位置でもって決定することができ、品質の高い音声合成を行うことが可能となる。

【0071】

また、圧縮すべき波形データの容量を必要最小限に抑制することができるので、解凍時間を短縮することができ、音声合成のリアルタイム性を確保することも可能となる。

## 【0072】

次に、本発明の実施の形態にかかる音声データ圧縮・解凍装置を実現するプログラムの処理の流れについて説明する。図8に本発明の実施の形態にかかる音声データ圧縮・解凍装置を実現するプログラムの処理の流れ図を示す。

## 【0073】

図8において、波形データを既存の音声波形辞書等から抽出して入力すると（ステップS81）、当該波形データの中で音声合成に使用する部分を特定し、その前後で圧縮する開始点及び終了点を仮設定する（ステップS82）。

## 【0074】

次に、仮設定した圧縮区間について圧縮及び解凍を行う（ステップS83）。解凍された波形データの品質が高ければ（ステップS84：Yes）、仮設定した圧縮区間を圧縮・解凍位置として確定し（ステップS85）、音声合成に用いる波形データとしてデータベースに登録する（ステップS86）。解凍された波形データの品質が高くなければ（ステップS84：No）、圧縮位置を再度仮設定して（ステップS87）、上述した処理を繰り返す。

## 【0075】

本発明の実施の形態にかかる音声データ圧縮・解凍装置を実現するプログラムを記憶した記録媒体は、図9に示す記録媒体の例に示すように、CD-ROM92-1やフロッピーディスク92-2等の可搬型記録媒体92だけでなく、通信回線の先に備えられた他の記憶装置91や、コンピュータ93のハードディスクやRAM等の記録媒体94のいずれでも良く、プログラム実行時には、プログラムはローディングされ、主メモリ上で実行される。

## 【0076】

また、本発明の実施の形態にかかる音声データ圧縮・解凍装置により生成された圧縮データ等を記録した記録媒体も、図9に示す記録媒体の例に示すように、CD-ROM92-1やフロッピーディスク92-2等の可搬型記録媒体92だけでなく、通信回線の先に備えられた他の記憶装置91や、コンピュータ93のハードディスクやRAM等の記録媒体94のいずれでも良く、例えば本発明にかかる音声データ圧縮・解凍装置を利用する際にコンピュータにより読み取られる

【 0 0 7 7 】

【発明の効果】

以上のように本発明にかかる音声データ圧縮・解凍装置によれば、波形データにおける圧縮位置や解凍位置を、音声合成時のS/N比が高い位置で決定することができ、品質の高い音声合成を行うことが可能となる。

【 0 0 7 8 】

また本発明にかかる音声データ圧縮・解凍装置によれば、圧縮すべき波形データの容量を必要最小限に抑制することができるので、解凍時間を短縮することができ、音声合成のリアルタイム性を確保することが可能となる。

【図面の簡単な説明】

【図 1】 従来の音声データ圧縮・解凍装置の構成図

【図 2】 本発明の実施の形態にかかる音声データ圧縮・解凍装置の構成図

【図 3】 本発明の他の実施例にかかる音声データ圧縮・解凍装置の構成図

【図 4】 本発明の他の実施例にかかる音声データ圧縮・解凍装置の構成図

【図 5】 本発明の実施の形態にかかる音声データ圧縮・解凍装置における音声合成の構成図

【図 6】 本発明の一実施例にかかる音声データ圧縮・解凍装置の構成図

【図 7】 本発明の他の実施例にかかる音声データ圧縮・解凍装置の構成図

【図 8】 本発明の実施の形態にかかる音声データ圧縮・解凍装置における処理の流れ図

【図 9】 記録媒体の例示図

【符号の説明】

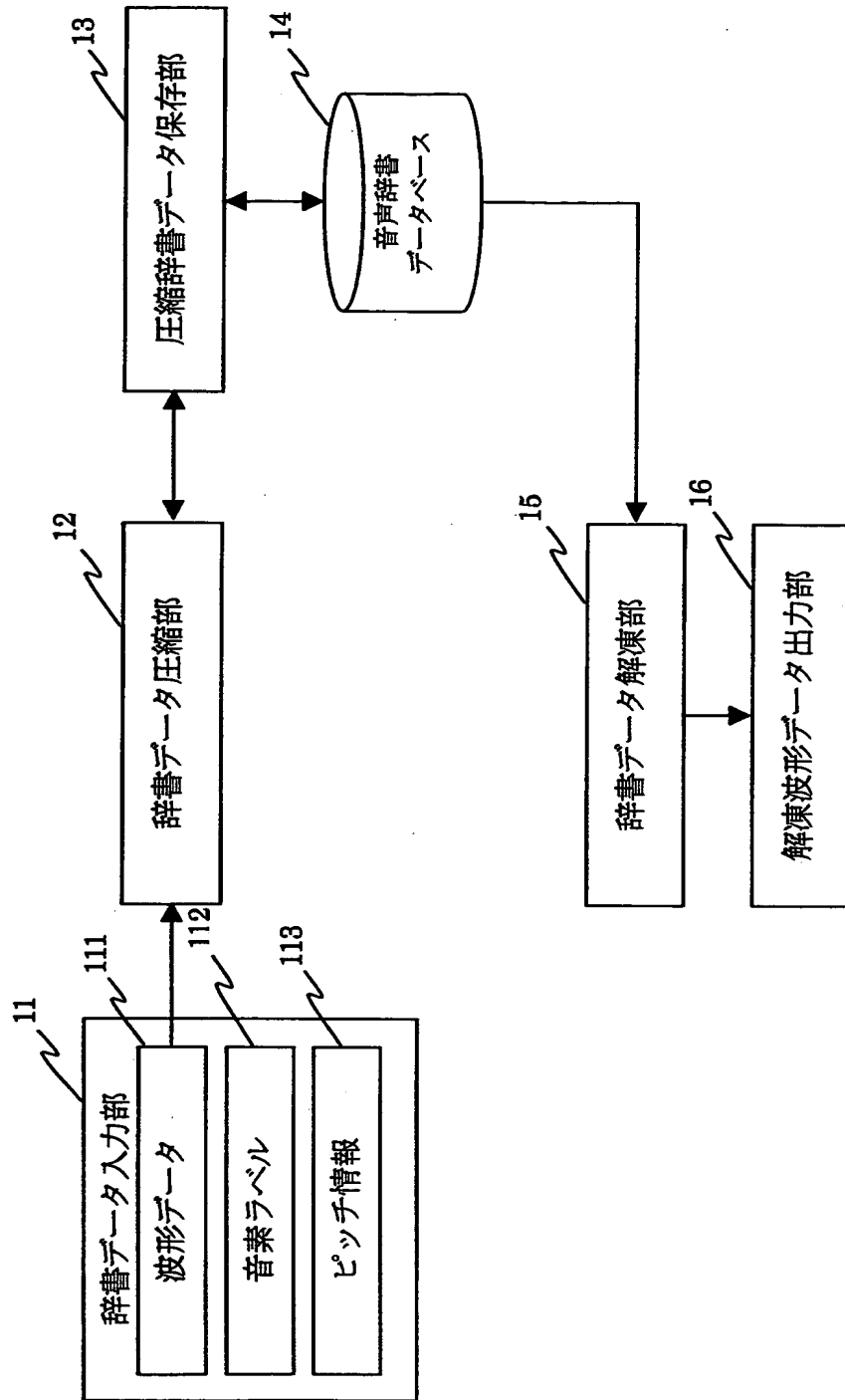
- 1 1 辞書データ入力部
- 1 2 辞書データ圧縮部
- 1 3、2 1 圧縮辞書データ保存部
- 1 4 音声辞書データベース
- 1 5 辞書データ解凍部
- 1 6 解凍波形データ出力部



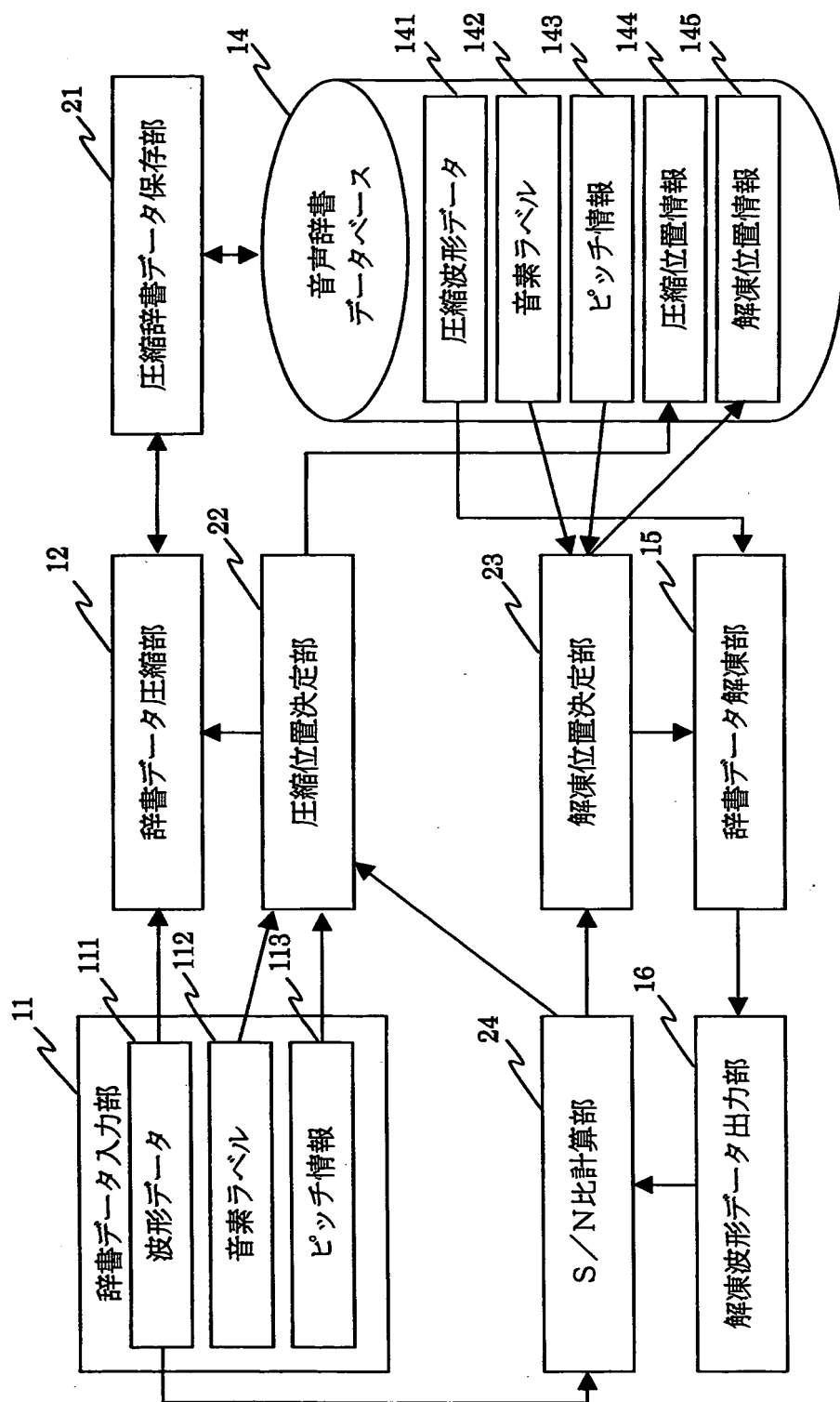
- 2 2 圧縮位置決定部
- 2 3 解凍位置決定部
- 2 4 S/N比計算部
- 5 1 音声合成部
- 9 1 回線先の記憶装置
- 9 2 CD-ROMやフロッピーディスク等の可搬型記録媒体
- 9 2-1 CD-ROM
- 9 2-2 フロッピーディスク
- 9 3 コンピュータ
- 9 4 コンピュータ上のRAM/ハードディスク等の記録媒体
- 1 1 1 波形データ
- 1 1 2、1 4 2 音素ラベル
- 1 1 3、1 4 3 ピッチ情報
- 1 4 1 圧縮波形データ
- 1 4 4 圧縮位置情報
- 1 4 5 解凍位置情報
- 2 2 1 無音区間削除部
- 2 2 2 有音区間波形生成部
- 2 2 3 圧縮区間設定部
- 2 2 4、2 3 1 音節抽出部
- 2 2 5 圧縮区間・S/N比保存部
- 2 3 2 音節波形区間抽出部
- 2 3 3 解凍区間設定部
- 2 3 4 解凍区間・S/N比保存部

【書類名】 図面

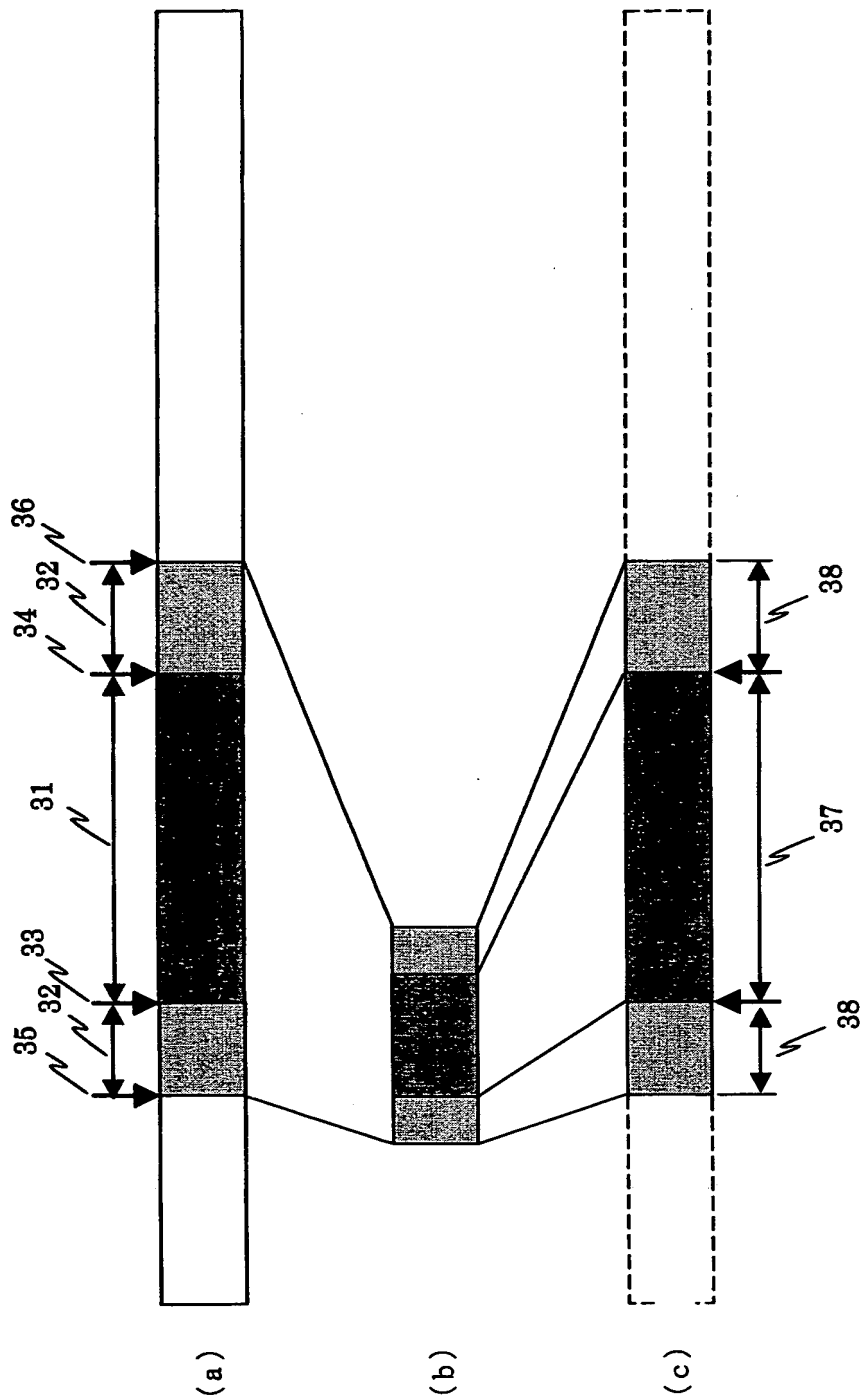
【図 1】



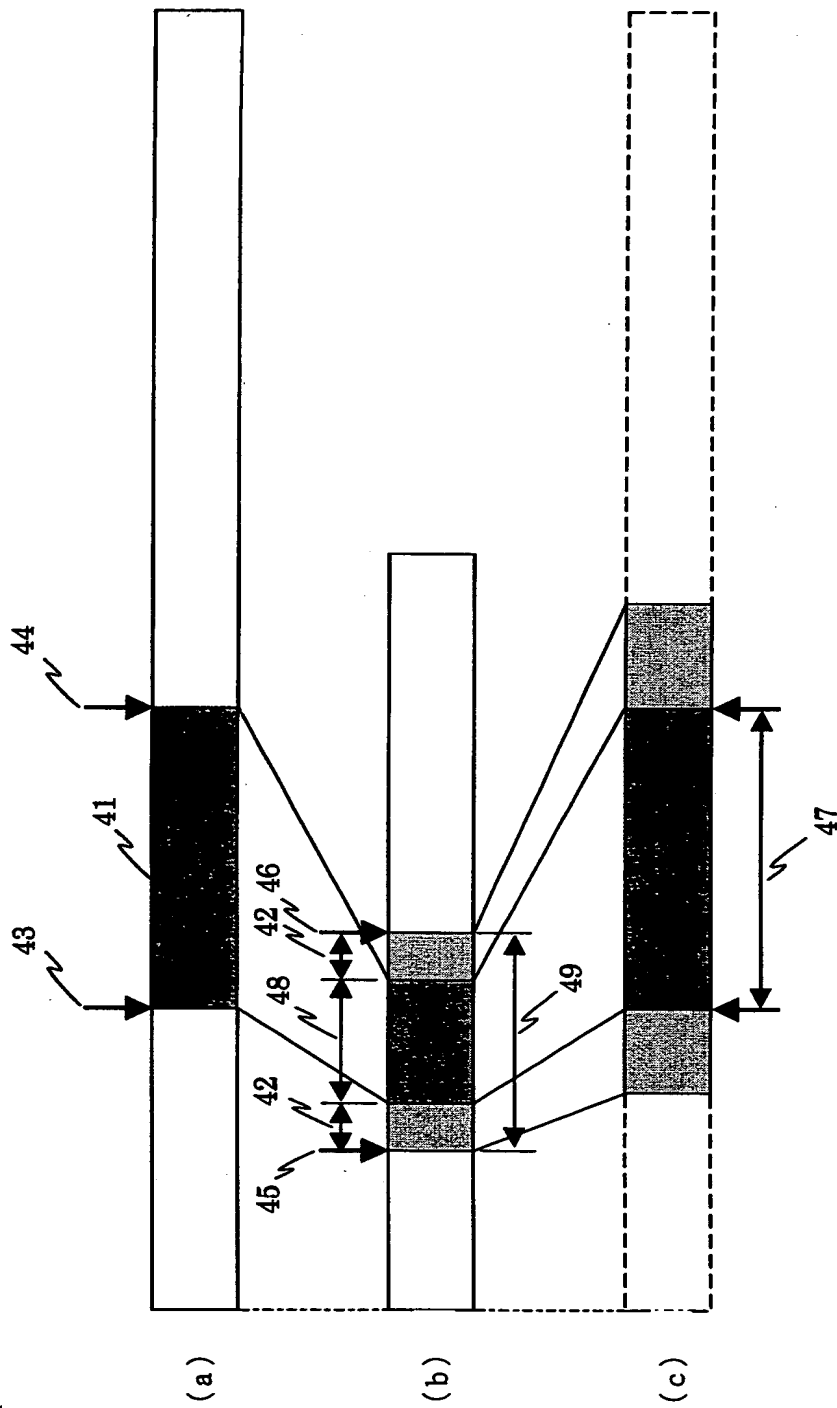
【図 2】



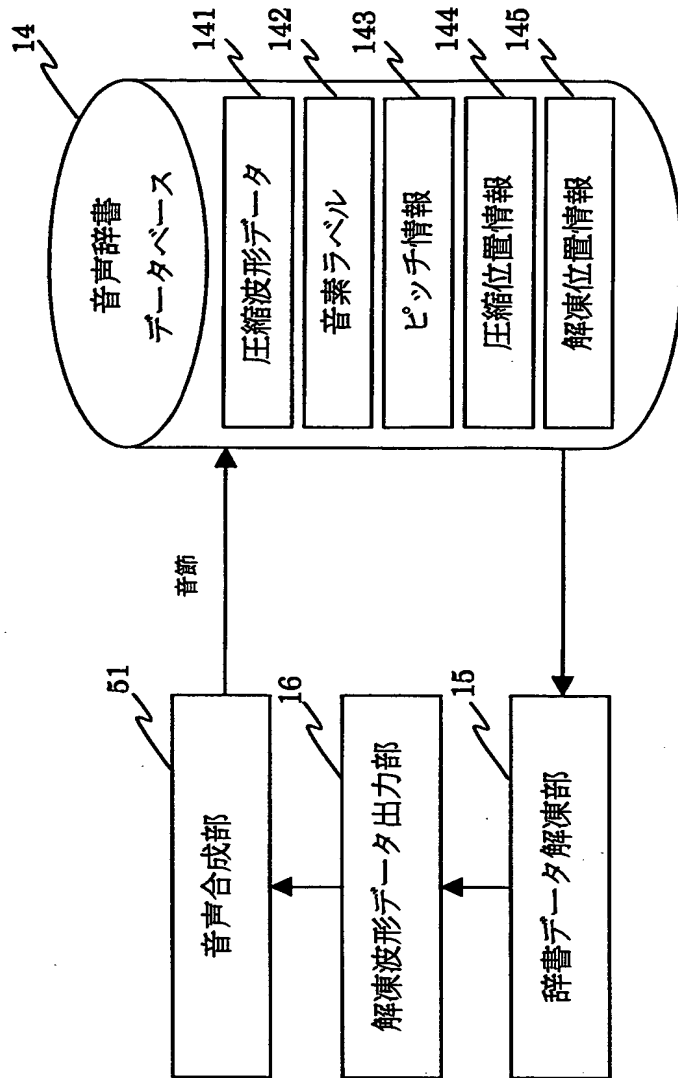
【図 3】



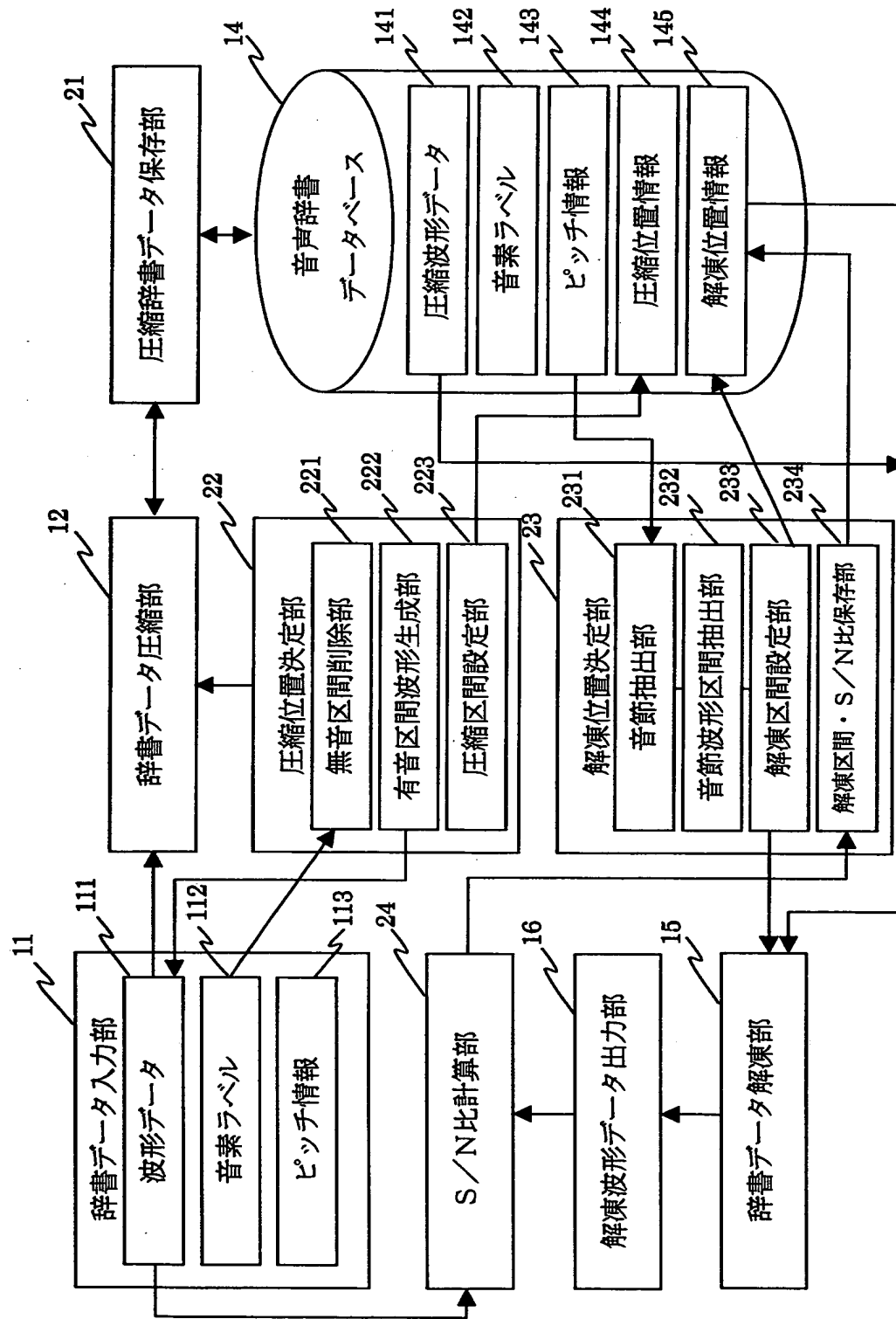
【図 4】



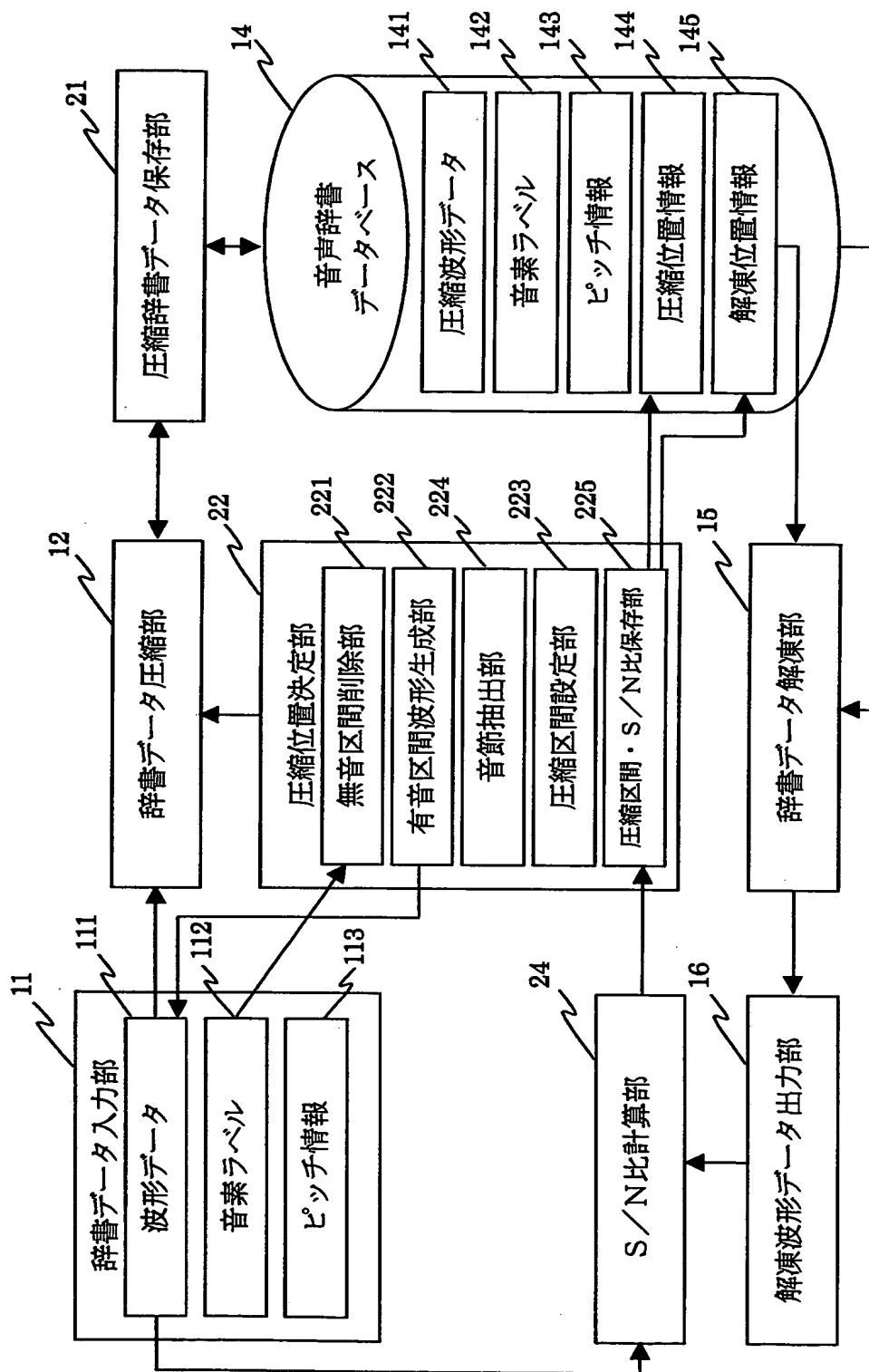
【図 5】



【図 6】

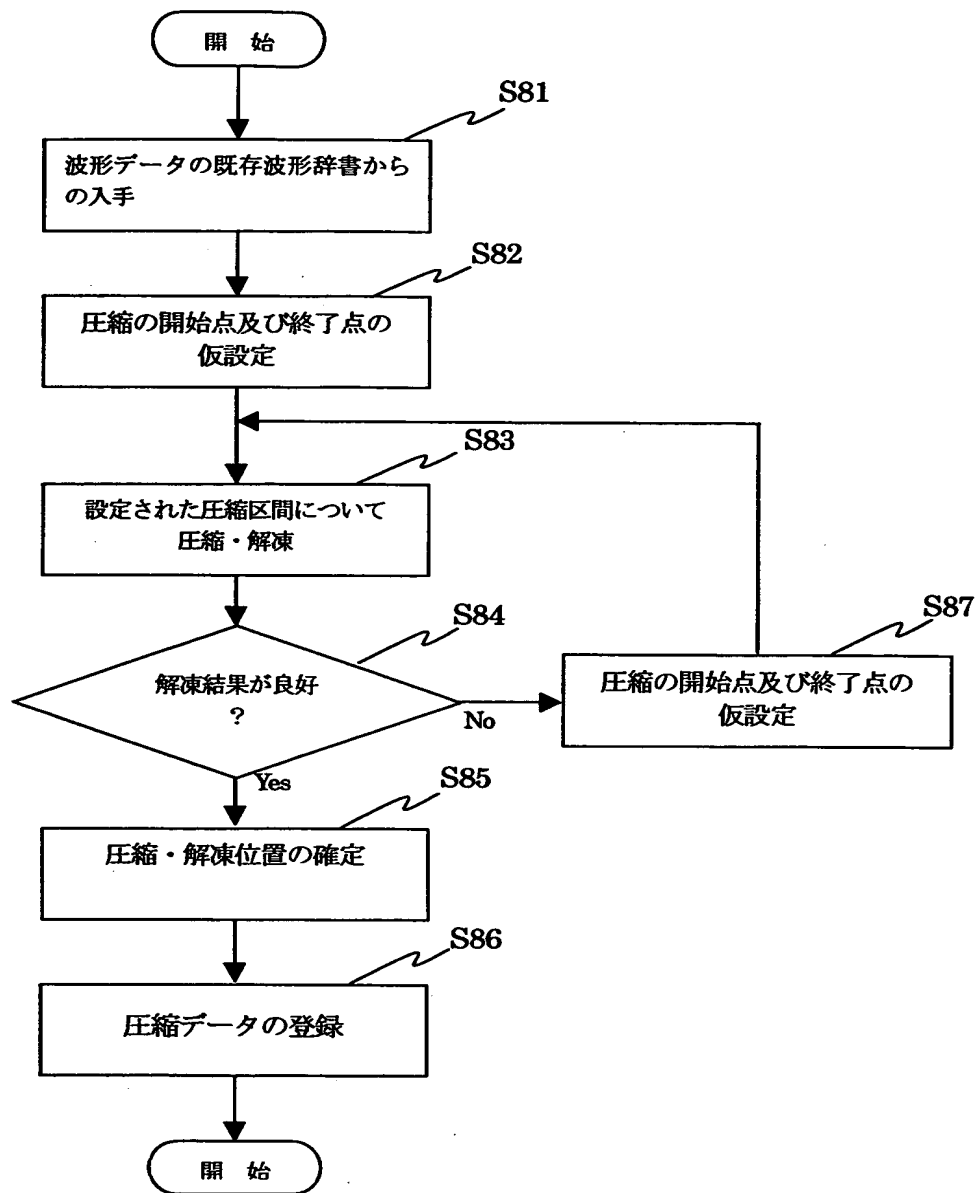


【図 7】

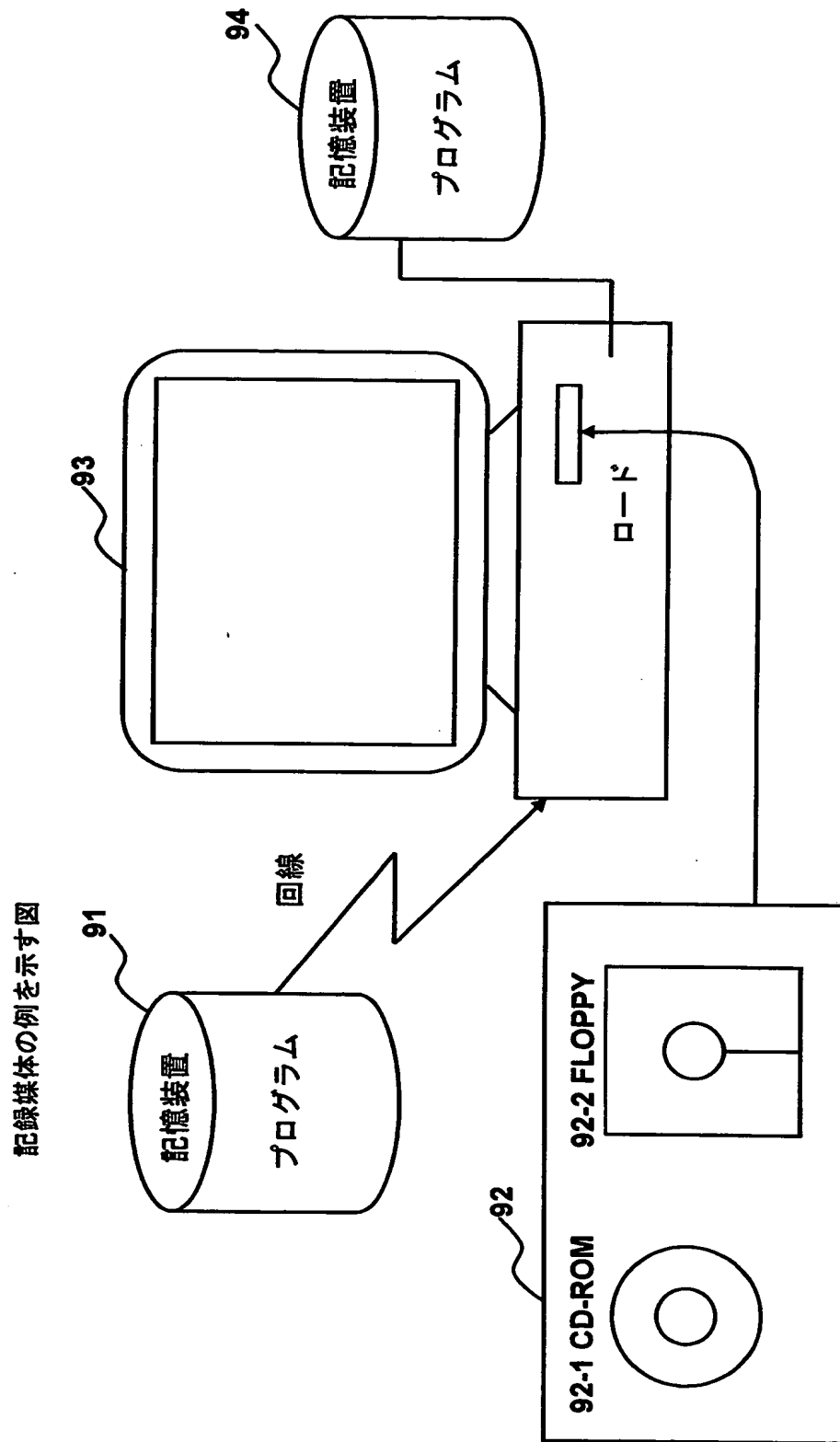




【図 8】



【図 9】



【書類名】 要約書

【要約】

【課題】 波形データにおける圧縮位置や解凍位置を是正して、音声合成のリアルタイム性を確保するとともに、高品質な音声合成を実現する音声データ圧縮・解凍装置及び方法を提供する。

【解決手段】 波形データを含む音声データを既存の音声波形辞書から抽出して入力し、波形データの中で音声合成に用いる部分を特定し、その前後で圧縮する開始点及び終了点を設定し、設定した圧縮区間について波形データを圧縮し、圧縮された波形データを解凍するとともに、圧縮された波形データの解凍結果の品質が最も高かった圧縮する開始点及び終了点で指定される圧縮区間を圧縮・解凍位置として確定し、音声合成に用いる波形データとして圧縮された波形データと圧縮する開始点及び終了点をデータベースに登録する。

【選択図】 図 2

出 願 人 履 歴 情 報

識別番号 [0 0 0 0 0 5 2 2 3]

1. 変更年月日	1 9 9 6 年 3 月 2 6 日
[変更理由]	住所変更
住 所	神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号
氏 名	富士通株式会社